An Assessment of the Accuracy of the RRIGS Hydration Potential: Comparison to Solutions of the Poisson–Boltzmann Equation

JOSEPH D. AUGSPURGER,* HAROLD A. SCHERAGA
Cornell University, Baker Laboratory of Chemistry, Ithaca, New York 14853-1301

Received 11 June 1996; accepted 28 October 1996

ABSTRACT: A rapid, pairwise hydration potential, the reduced radius independent Gaussian sphere (RRIGS) approximation, has been presented recently. Because experimental values of the conformational dependence of the hydration free energy are unavailable, this hydration potential is testable by comparison to a presumably more accurate (yet more computationally intensive) model. One such method is the electrostatic hydration approach, which models the protein as a collection of point charges in a low-dielectric medium and the solvent as a high-dielectric continuum. The electrostatic free energy can be determined by solving the Poisson–Boltzmann equation, which is carried out with the program DelPhi. The total free energy of hydration is calculated by adding a free energy of cavity formation term to this electrostatic term. Comparison is made for many conformations of two proteins, bovine pancreatic trypsin inhibitor (BPTI) and the carboxy-terminal fragment of the L7/L12 ribosomal protein (CTF). Thirty-nine near-native structures of BPTI, previously generated by Ripoll and coworkers, and 150 conformations of CTF, generated by a threading algorithm to cover a wide range of conformational space, were used in these comparisons. It is shown that, for the neutral forms of these proteins, the RRIGS hydration potential correlates very well with the electrostatic model hydration free energy, although the correlation is better for the CTF

* Special Fellow of the Leukemia Society of America.
Correspondence to: H. A. Scheraga
Contract grant sponsor: National Institutes of Health; contract grant number A: GM-14312
Contract grant sponsor: National Science Foundation; contract grant number: MCB95-13167
Contract grant sponsor: National Foundation for Cancer Research
Contract grant sponsor: Association for International Cancer Research
Contract grant sponsor: Leukemia Society of America

 CCC 0192-8651 / 97 / 081072-07

conformations than for the near-native BPTI conformations. For charged forms, the correlation is much poorer. These results serve as evidence that solvent-exposure models of hydration, which leave out cooperative effects between different groups, may be appropriate for modeling neutral or slightly charged species, because these cooperative effects are likely to be small. However, for highly charged species where cooperative effects are surely large, such an approach will be less accurate. © 1997 by John Wiley & Sons, Inc. *J Comput Chem* **18**: 1072–1078, 1997

Introduction

One approach to the prediction of protein structure is minimization of a conformationally dependent energy function. This approach is based on the widely accepted notion that protein structures are thermodynamically stable; that is, that they adopt the structure which corresponds to the global minimum of the conformational energy hypersurface. The successful application of this approach to protein structure prediction requires that the energy function be computationally efficient (so that the search of the energy hypersurface can be accomplished in a reasonable time) and that it accurately reflect the energetics of the protein in solution as a function of its conformation.

We have recently introduced a method to model the conformationally dependent free energy of hydration that satisfies the first of these two criteria. The reduced radius, independent Gaussian sphere (RRIGS) hydration potential was developed[1] for use with the empirical conformational energy program for peptides (ECEPP),[2–5] which uses fixed bond lengths and bond angles. It was shown to be quite rapid, because it is based on a pairwise calculation of the exposed volume of the hydration shell (VHS) about each atom. RRIGS is based on the approximation that the total hydration free energy is a sum of the hydration free energies of the individual atoms, and that the hydration free energy of each atom is proportional to its solvent exposure (as measured by the exposed VHS). The VHS is related to the hydration free energy of different atom types by empirically determined parameters. The present work assesses the accuracy of the RRIGS potential, that is, the aforementioned second criterion.

Ideally, a direct comparison should be made to experimental data to assess the accuracy of a theoretical model. In this case, there are no direct, experimental data for conformationally dependent hydration free energies of proteins. Therefore, we chose to evaluate RRIGS by comparison to a widely used and presumably more accurate (although more computationally intensive) method of calculating conformationally dependent hydration free energies. The electrostatic model of hydration of a protein treats the protein as a collection of point charges in a low-dielectric medium, immersed in a high-dielectric continuum; the electrostatic free energy is calculated by solving the Poisson–Boltzmann equation. This is augmented by a term that reflects the free energy of cavity formation in water. Whereas this method is widely used, it is computationally too intensive to be appropriate with currently available computers to carry out a search for the global minimum of the very complex energy hypersurface of a protein. Thus, we compare the hydration free energy predicted by the rapid RRIGS potential to those of the electrostatic hydration model to assess its accuracy.

Theory and Methods

To make a significant test of the RRIGS hydration potential requires that many conformations which span a large portion of conformational space be examined. To generate such an ensemble of structures, a threading algorithm was used. In this threading algorithm, the amino acid sequence of a given protein is made to adopt the structures of many different proteins. Obviously, the sequence to be threaded must possess fewer residues than the target structures into which it will be threaded; furthermore, the side-chain conformations cannot be defined as they will be (mostly) different. The presence of disulfide bonds complicates threading, because the individual cysteine residues of the sequence to be threaded are not likely to be close to each other in the target structures. Thus, we choose to apply this threading approach to generate an ensemble of structures of the carboxy-terminal fragment (CTF) of the L7/L12 ribosomal protein; this fragment contains 68 residues (53–120),

possesses no disulfide bonds, and a high-resolution crystal structure is known.[6]

A data base of 197 nonhomologous proteins was taken from the protein data bank for which high-resolution structures were available (resolution < 2.0 Å). For each of 150 structures taken from this data base, both the structure into which the amino acid sequence of CTF would be threaded as well as the starting position (residue number) were chosen randomly. The following protocol was used to locate a low-energy local minimum of the ECEPP + RRIGS conformational energy function for CTF which was close to the threaded protein structure:

1. A terminally blocked, 68-residue poly-L-alanine backbone structure was generated by starting from the dihedral angles, ϕ, ψ, ω, of the experimental target structure and minimizing the rms deviation between the backbone atoms of poly-L-alanine and the backbone atoms of the target structure. This typically resulted in an rms deviation from the target structure for the backbone atoms of 0.3–0.6 Å.
2. Starting with these resulting backbone dihedral angles, ϕ, ψ, ω, for CTF, each side-chain dihedral angle of CTF was minimized sequentially while the backbone dihedral angles were held fixed. Three cycles of minimization of all side-chain dihedral angles were carried out.
3. Finally, from the resulting minimized side-chain structure, a local minimization with all backbone and side-chain dihedral angles allowed to vary was carried out by means of the SUMSL minimization routine[7] to arrive at the final, local minimum conformation.

When this protocol was applied to the experimental CTF structure, the resulting structure possessed a backbone rms deviation of 0.97 Å from the experimental structure and an ECEPP/3[5] energy of -413.3 kcal/mol. In a small number of cases, unphysical structures (e.g., with a side chain overlapping the backbone) resulted when one of the CTF prolines (which in ECEPP have fixed dihedral angles, ϕ) was positioned just before a turn in the target structure, and such high-energy conformations were discarded. This threading algorithm was parallelized to take advantage of the IBM SP2 parallel architecture, using up to 32 processors in one run, and was used to generate 150 different CTF conformations.

A second set of structures was also examined. This consisted of 39 structures of BPTI which are local minima of the ECEPP function. They were generated by application of the self-consistent electrostatic field (SCEF) and electrostatically driven Monte Carlo (EDMC) methods starting from a near-native conformation; the largest rms deviation of the backbone atoms of these conformations from one of the crystal structures was about 2 Å.[8] These represent a different type of ensemble of conformations, namely a cluster of closely related structures. They were included in this assessment to test whether the RRIGS potential will accurately discriminate among native and near-native structures in terms of the hydration free energy.

The RRIGS hydration potential is defined to be:

$$\Delta G_{\mathrm{hyd}}^{\mathrm{RIGGS}} = \sum_i \delta_i (\mathrm{VHS})_i \qquad (1)$$

where i represents all atoms in the protein (except for the nonpolar hydrogens because a united-atom approach was used to treat the hydration parameters[1]) and δ_i is an empirically determined parameter. A set of δ_i parameters, appropriate for modeling all 20 naturally occurring amino acids in their neutral states, using 24 different atom types, was reported previously.[1] These parameters were determined by a least squares fit of eq. (1) to 140 experimentally determined free energies of hydration for small organic molecules. An enlarged data set was generated by including nine additional (ionic) species, for which experimentally derived values of ΔG_{hyd} have been reported.[9] Two new parameter sets were generated, in which four and six, respectively, new atom types were added to describe the charged species, by least squares fitting.

The free energy of hydration based on the electrostatic model can be described by:

$$\Delta G_{\mathrm{hyd}}^{\mathrm{ELEC}} = \Delta G_{\mathrm{PB}} + \Delta G_{\mathrm{cav}} \qquad (2)$$

where ΔG_{PB} represents the electrostatic free energy found by solving the Poisson–Boltzmann equation and ΔG_{cav} is the free energy of cavity formation. ΔG_{PB} was calculated by means of the DelPhi program, version 5.0.[10] The inner and outer dielectric constants were set to 2.0 and 80.0, respectively. The number of grid elements was determined by setting the box filled parameter to 80% with a grid size of 0.5 Å. Convergence was tested by using a smaller grid size (0.33 Å) as well as by adding a slight offset to the grid origin; neither of these led to changes of more than 0.5%

in the values of ΔG_{PB}. Two sets of atomic charges were examined. The PARSE set of charges[11] had been determined empirically to reproduce experimental free energies of hydration, and AMBER charges (from the AMBER force field) had been generated by fitting point charge models to electrostatic potential maps calculated by *ab initio* methods.[12]

The cavity formation term of Sitkoff et al.[11] is given by

$$\Delta G_{cav} = \gamma S_{acc} + b \qquad (3)$$

where $\gamma = 0.005$ kcal/mol/Å^2, S_{acc} is the total solvent accessible surface area, and $b = 0.86$ kcal/mol; the γ and b parameters had been determined by fitting experimental alkane transfer free energies to the computed accessible surface area.[11] The molecular surface program of Connolly[13,14] was used to calculate the surface area (S_{acc}) of the different protein conformations.

Results and Discussion

Figure 1 depicts a comparison between ΔG_{hyd}^{RRIGS} and ΔG_{hyd}^{ELEC} for the 39 BPTI structures and the 150 CTF structures for both the PARSE and AMBER charge sets, with both proteins in the neutral state. As can be seen from the figure, there is a strong correlation between the RRIGS hydration free energies and the electrostatic hydration free energies for both charge sets, with correlation coefficients of 0.904 and 0.908 with the PARSE and AMBER charge sets, respectively. The ΔG_{hyd}^{RRIGS} values are seen to match the ΔG_{hyd}^{ELEC} (PARSE) results more closely. This result is not unexpected because both the RRIGS potential and the PARSE charge set were developed empirically to reproduce experimental hydration free energies, whereas the AMBER charge set was not.

These results for CTF (depicted in Fig. 1) show that, for a wide range of conformational space which spans a range of hydration free energies of several hundred kilocalories per mole, the RRIGS hydration model correlates strongly with the electrostatic hydration model. This is encouraging, indicating that, for widely different conformations of neutral species, the RRIGS hydration potential can provide essentially the same information as the electrostatic hydration model, but much faster in terms of computational time. On a single thin node of an IBM SP2 computer (the equivalent of an RS6000/390), DelPhi calculations required about

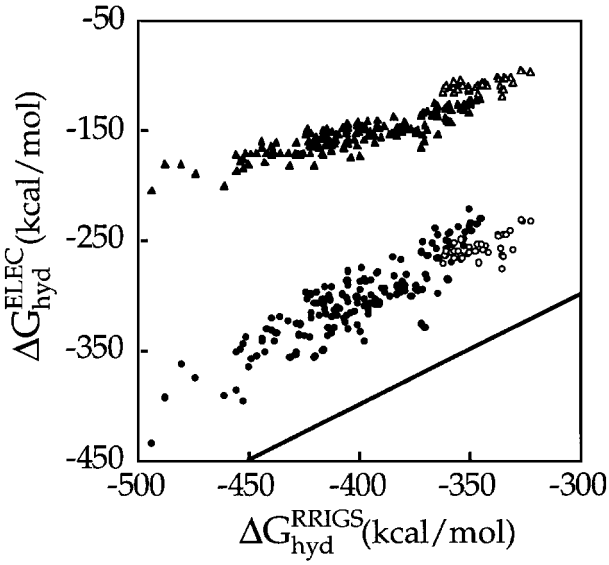

FIGURE 1. The free energy of hydration of 150 conformations of CTF and 39 conformations of BPTI as predicted by the electrostatic model (ΔG_{hyd}^{ELEC}) and by the RRIGS potential (ΔG_{hyd}^{RRIGS}) for both proteins in their neutral states are compared for both the PARSE charge set and the AMBER charge set. Filled circles (●) represent CTF with PARSE charges, open circles (○) BPTI with PARSE charges, filled triangles (▲) CTF with AMBER charges, and open triangles (△) BPTI with AMBER charges. The solid line would represent exact agreement ($\Delta G_{hyd}^{ELEC} = \Delta G^{RRIGS}{}_{hyd}$).

80 seconds for CTF and about 40 seconds for BPTI, whereas a single RRIGS evaluation required only 0.8 and 0.6 seconds, respectively. This represents a speedup of almost two orders of magnitude by the RRIGS hydration potential compared to DelPhi.

Close inspection of Figure 1 reveals that, for the BPTI conformations which are a set of near-native conformations, there are more discrepancies in the relative ordering of the hydration free energy as calculated by the RRIGS hydration potential and by DelPhi. The conformation that possesses the lowest electrostatic hydration free energy has a RRIGS hydration energy about 50 kcal/mol higher than the lowest RRIGS hydration energy. This indicates that the RRIGS hydration potential may be most suited for rapid screening of large numbers of widely varying conformations, as may be encountered and examined with a low-resolution conformational energy function.

However, the neutral charge state for these two proteins is not a physically realistic one; at pH 7, the GLU, ASP, LYS, and ARG side chains would all be expected to be ionized. A more realistic test would involve the examination of the two proteins

in their charged states. Thus, we now compare the RRIGS potential to the electrostatic hydration model for the fully charged states of BPTI (+5) and CTF (+2). We first examined the charged states using parameter sets generated by fitting experimental free energies of hydration of small molecules, including the few known ionic species.[9] The results (not shown) exhibit essentially no correlation between ΔG_{hyd}^{RRIGS} and ΔG_{hyd}^{ELEC}.

A difficulty in this approach is the dearth of experimental data for charged species, leading to a great imbalance between the number of neutral (140) and charged (9) species in the data set used to parameterize the hydration free energies. Because of this lack of experimental data, we chose to use the data from the electrostatic hydration model in place of experimental hydration free energies to make a different kind of test expressed in terms of the following question: Can the ΔG_{hyd}^{ELEC} values for a training subset from the 150 CTF and 39 BPTI test conformations be used to generate a set of $\boldsymbol{\delta}_i$ values which would allow ΔG_{hyd}^{RRIGS} to reproduce Δ_{hyd}^{ELEC} for the remaining conformations?

First, 50 CTF conformations and all 39 BPTI conformations were used as a training set to generate a set of $\boldsymbol{\delta}_i$ values, which were then used to calculate ΔG_{hyd}^{RRIGS} for the remaining 100 CTF conformations. This result is shown in Figure 2. The correlation between ΔG_{hyd}^{RRIGS} and ΔG_{hyd}^{ELEC} is much less than the results for neutral proteins. The resulting correlation coefficients were 0.375, −0.040, and 0.118 for the BPTI conformations, the 50 CTF conformations included in the training set, and the 100 test CTF conformations, respectively.

A second fitting was carried out using as the training set the 100 CTF conformations and 20 BPTI conformations. The results for the fitting of 50 conformations of CTF and 19 conformations of BPTI with this training set are given in Figure 3. The CTF results are inaccurate, as in the first fitting shown in Figure 2; moreover, the BPTI conformations now exhibit a negative correlation. The resulting correlation coefficients were 0.455, 0.322, −0.731, and 0.173, for the BPTI and CTF conformations in the training set and the BPTI and CTF test conformations, respectively. These results indicate that the RRIGS hydration method cannot reproduce the hydration energies of the electrostatic hydration model for the proteins in these highly charged states.

To try to identify the inadequacy of the RRIGS hydration potential in charged proteins, two more

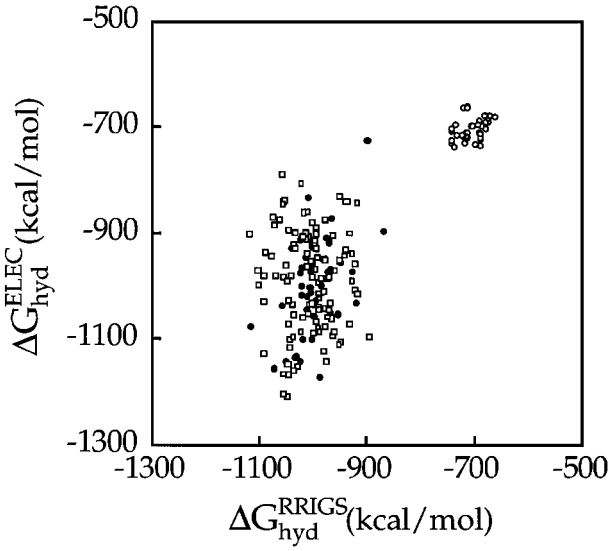

FIGURE 2. Comparison of the free energy of hydration for 100 charged CTF conformations, indicated by open squares (□), as predicted by the electrostatic model (ΔG_{hyd}^{ELEC}) using PARSE charges, and by the RRIGS potential (ΔG_{hyd}^{RRIGS}), where the $\boldsymbol{\delta}_i$ parameters were determined by fitting the ΔG_{hyd}^{ELEC} values from a training set of 50 CTF conformations and 39 BPTI conformations. The fitted values for the training set are also depicted, with filled circles (●) indicating the CTF and open circles (○) the BPTI conformations.

fitting procedures were carried out, using all of the conformations but treating BPTI and CTF separately. These results are shown in Figures 4 and 5, respectively, and the correlation coefficients are 0.947 for BPTI and 0.455 for CTF. There is a significant difference between the capability of the RRIGS method to reproduce the electrostatic hydration energies for these two different data sets; the fit is much better for the BPTI than for the CTF structures.

It is important to reiterate the difference between these two sets of conformations. The CTF set contains widely different conformations, with the electrostatic hydration varying over a range of nearly 500 kcal/mol. However, the BPTI set contains *near-native* conformations, with the largest backbone rms deviation from the native structure being 2 Å or less. A significant difference between the RRIGS method and the electrostatic method is that the RRIGS approach neglects cooperative effects; it is designed to be pairwise. These cooperative effects, however, are included in the electrostatic method. Hence, it appears that the RRIGS method can reproduce ΔG_{hyd}^{ELEC} for neutral proteins

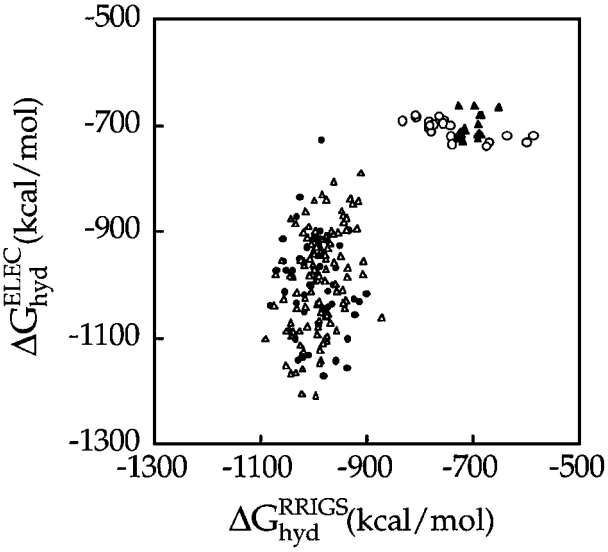

FIGURE 3. Comparison of the free energy of hydration for 50 charged CTF conformations, indicated by filled circles (●), and 19 charged BPTI conformations, indicated by open circles (○), as predicted by the electrostatic model (ΔG_{hyd}^{ELEC}) using PARSE charges, and by the RRIGS potential (ΔG_{hyd}^{RRIGS}), where the δ_i parameters were determined by fitting the ΔG_{hyd}^{ELEC} values from a training set of 100 CTF conformations and 20 BPTI conformations. The fitted values for the training set are also depicted, with open triangles (△) indicating the CTF and filled triangles (▲) the BPTI conformations.

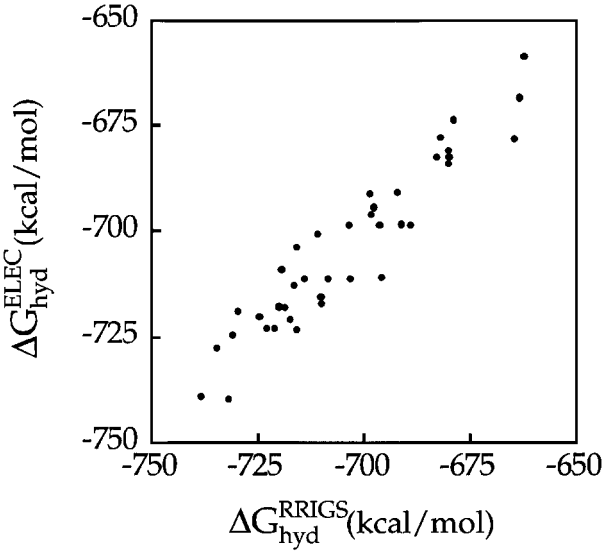

FIGURE 4. Comparison of ΔG_{hyd}^{RRIGS} to ΔG_{hyd}^{ELEC} for all 39 BPTI conformations when the δ_i parameters were generated by fitting the 39 ΔG_{hyd}^{ELEC} values of BPTI, using the RRIGS algorithm.

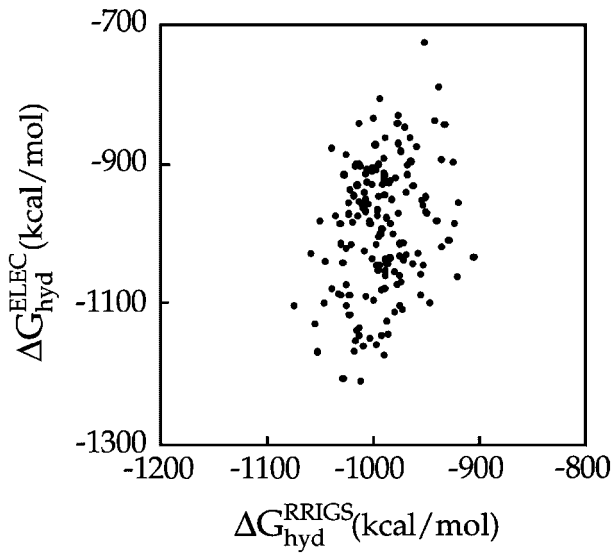

FIGURE 5. Comparison of ΔG_{hyd}^{RRIGS} to ΔG_{hyd}^{ELEC} for all 150 CTF conformations when the δ_i parameters were generated by fitting the 150 ΔG_{hyd}^{ELEC} values of CTF, using the RRIGS algorithm.

(Fig. 1) in which these cooperative effects would be expected to be smaller than for charged proteins.

As seen in Figure 4, the RRIGS method also appears to reproduce ΔG_{hyd}^{ELEC} for charged BPTI. A possible reasons for this is that, because the conformations are roughly similar but with an energy range of about 100 kcal/mol, the cooperative effects are likely to be nearly constant. On the other hand, for the charged CTF conformations, where these cooperative effects are large and highly variable over the wide range of structures examined, a pairwise approach is simply inadequate (Fig. 5).

This conjecture regarding the importance of cooperative effects on the ionizable side-chain groups receives support from the recent study of a 17-residue peptide by Ripoll et al.[15] They used a new method to compute the helix content of this peptide as a function of pH, whereby the degree of ionization of each ionizable group was calculated for different conformations. These calculations showed that the degree of ionization of the different ionizable side chains is strongly influenced by the conformation of the peptide, which in turn strongly influences the hydration energy. These results,[15] together with those of the present study, illustrate the necessity of including cooperative effects to model the interaction of highly charged proteins with water accurately.

Conclusion

The conformational hydration free energy of a large number of conformations of two proteins, as predicted by the RRIGS pairwise potential, has been compared to the more computationally intensive electrostatic hydration model. It has been shown to correlate quite strongly to the hydration free energies of the electrostatic hydration model for a neutral protein in a large sampling of conformational space, somewhat less so for a sampling of near-native conformations, and for roughly similar conformations of a charged protein (even as the hydration energy varies by nearly 100 kcal/mol). However, for very different conformations of a charged protein, it seems inadequate to model the hydration free energy accurately. We attribute this to the importance of cooperative electrostatic effects in this case. These results indicate that the rapid RRIGS potential appears to be well suited for studying neutral or slightly charged peptides in solution.

Acknowledgments

All computations reported in this study were carried out on the IBM SP2 computer at the Cornell National Supercomputing Facility, a resource of the Cornell Center for Theory and Simulation in Science and Engineering which receives major funding from the National Science Foundation, the State of New York, the IBM Corporation, and members of its Corporate Research Institute, with additional funds from the National Institutes of Health. We thank Professor Barry Honig for providing the most recent version of DelPhi, and Daniel R. Ripoll for the ECEPPAK version of the ECEPP/3 program, which was modified to carry out the threading algorithm.

References

1. J. D. Augspurger and H. A. Scheraga, *J. Comput. Chem.* **17**, 1549 (1996).
2. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.* **79**, 2361 (1975).
3. G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.* **87**, 1883 (1983).
4. M. J. Sippl, G. Némethy, and H. A. Scherga, *J. Phys. Chem.* **88**, 6231 (1984).
5. G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, *J. Phys. Chem.* **96**, 6472 (1992).
6. M. Leijonmarck and A. Liljas, *J. Mol. Biol.* **195**, 555 (1987).
7. D. M. Gay, *ACM Trans. Math Software* **9**, 503 (1983).
8. D. R. Ripoll, L. Piela, M. Vásquez, and H. A. Scheraga, *Proteins* **10**, 188 (1991).
9. Y. K. Kang, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.* **91**, 4118 (1987).
10. A. Nicholls, K. A. Sharp, and B. Honig, *DelPhi*, Department of Biochemistry and Molecular Biology, Columbia University, New York, 1990.
11. D. Sitkoff, K. A. Sharp, and B. Honig, *J. Phys. Chem.* **98**, 1978.
12. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.* **7**, 230 (1986).
13. M. L. Connolly, *J. Appl. Cryst.* **16**, 548 (1983).
14. M. L. Connolly, *J. Am. Chem. Soc.* **107**, 1118 (1985).
15. D. R. Ripoll, Y. N. Vorobjev, A. Liwo, J. A. Vila, and H. A. Scheraga, *J. Mol. Biol.* **264**, 770 (1996).